



---

## An Algebraic Approach to Inference in Complex Networked Structures

Jose Moura  
CARNEGIE MELLON UNIVERSITY

---

07/09/2015  
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory  
AF Office Of Scientific Research (AFOSR)/ RTA2  
Arlington, Virginia 22203  
Air Force Materiel Command

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Executive Services, Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p><b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.</b></p>				
<b>1. REPORT DATE (DD-MM-YYYY)</b> 02-09-2015		<b>2. REPORT TYPE</b> Final Performance		<b>3. DATES COVERED (From - To)</b> 01-04-2012 to 31-03-2015
<b>4. TITLE AND SUBTITLE</b> An Algebraic Approach to Inference in Complex Networked Structures			<b>5a. CONTRACT NUMBER</b>	
			<b>5b. GRANT NUMBER</b> FA9550-12-1-0087	
			<b>5c. PROGRAM ELEMENT NUMBER</b> 61102F	
<b>6. AUTHOR(S)</b> Jose Moura			<b>5d. PROJECT NUMBER</b>	
			<b>5e. TASK NUMBER</b>	
			<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> CARNEGIE MELLON UNIVERSITY 5000 FORBES AVENUE PITTSBURGH, PA 15213-3815 US			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> AF Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> AFRL/AFOSR RTA2	
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> A DISTRIBUTION UNLIMITED: PB Public Release				
<b>13. SUPPLEMENTARY NOTES</b>				
<b>14. ABSTRACT</b> <p>Analysis and processing of very large data sets, or big data, poses a significant challenge. Massive data sets are collected and studied in numerous domains, from engineering sciences to social networks, biomolecular research, commerce, and security. Extracting valuable information from big data requires innovative approaches that efficiently process large amounts of data and utilize their structure. This research project developed a paradigm for large-scale data analysis based on the discrete signal processing (DSP) on graphs (DSPG). DSPG extends signal processing concepts and methodologies from the classical signal processing theory to data indexed by general graphs. We introduced fundamental concepts of DSPG, including graph signals and graph filters, graph Fourier transform, graph frequency, and spectrum ordering that extended their counterparts from classical signal processing theory. Big data analysis presents several challenges to DSPG, in particular, in filtering and frequency analysis of very large data sets. We showed how to analyze these large data sets by considering product graphs as a graph model that helps extend the application of DSPG methods to large data sets through efficient implementations based on parallelization and vectorization. We illustrated the applicability of DSPG with numerous studies that are relevant in applications</p>				
<b>15. SUBJECT TERMS</b> <p>Network Characterization, Heterogeneous Systems</p>				
<b>16. SECURITY CLASSIFICATION OF:</b>				

<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified	<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> Jose Moura
					<b>19b. TELEPHONE NUMBER</b> <i>(Include area code)</i> 412-268-6341

# Final Report: An Algebraic Approach to Inference in Complex Networked Structures

FA9550-12-1-0087

4/1/2012-3/31/2015

José M F Moura

Carnegie Mellon University

Dep. of Electrical and Computer Eng., CMU, 5000 Forbes Av, Pittsburgh, PA 15213

moura@ece.cmu.edu

THERE is an explosion of interest in processing and analyzing large datasets collected in very different settings, including social and economic networks, information networks, internet and the world wide web, immunization and epidemiology networks, molecular and gene regulatory networks, citation and coauthorship studies, friendship networks, as well as physical infrastructure networks like sensor networks, power grids, transportation networks, and other networked critical infrastructures. We briefly contrast our approach in this project with existing work.

## A. Brief review of the literature

Many authors focus on the underlying *relational structure* of the data by: 1) inferring the structure from community relations and friendships, or from perceived alliances between agents as abstracted through game theoretic models [1], [2]; 2) quantifying the connectedness of the world; and 3) determining the relevance of particular agents, or studying the strength of their interactions. Other authors are interested in the *network function* by quantifying the impact of the network structure on the diffusion of disease, spread of news and information, voting trends, imitation and social influence, crowd behavior, failure propagation, global behaviors developing from seemingly random local interactions [2], [3], [4]. Much of these works either develop or assume network models that capture the interdependencies among the data and then analyze the structural properties of these networks. Models often considered may be deterministic like complete or regular graphs, or random like the Erdős-Rényi and Poisson graphs, the configuration and expected degree models, small world or scale free networks [2], [4], to mention a few. These models are used to quantify network characteristics, such as connectedness, existence and size of the giant component, distribution of component sizes, degree and clique distributions, and node or edge specific parameters including clustering coefficients, path length, diameter, betweenness and closeness centralities.

Another body of literature is concerned with inference and learning from such large datasets. Much work falls under the generic label of graphical models [5], [6], [7], [8], [9], [10]. In graphical models, data is viewed as a family of random variables indexed by the nodes of a graph, where the graph captures

probabilistic dependencies among data elements. The random variables are described by a family of joint probability distributions. For example, directed (acyclic) graphs [11], [12] represent Bayesian networks where each random variable is independent of others given the variables defined on its parent nodes. Undirected graphical models, also referred to as Markov random fields [13], [14], describe data where the variables defined on two sets of nodes separated by a boundary set of nodes are statistically independent given the variables on the boundary set. A key tool in graphical models is the Hammersley-Clifford theorem [13], [15], [16], and the Markov-Gibbs equivalence that, under appropriate positivity conditions, factors the joint distribution of the graphical model as a product of potentials defined on the cliques of the graph. Graphical models exploit this factorization and the structure of the indexing graph to develop efficient algorithms for inference by controlling their computational cost. Inference in graphical models is generally defined as finding from the joint distributions lower order marginal distributions, likelihoods, modes, and other moments of individual variables or their subsets. Common inference algorithms include belief propagation and its generalizations, as well as other message passing algorithms. A recent block-graph algorithm for fast approximate inference, in which the nodes are non-overlapping clusters of nodes from the original graph, is in [17]. Graphical models are employed in many areas; for sample applications, see [18] and references therein.

Extensive work is dedicated to discovering efficient data representations for large high-dimensional data [19], [20], [21], [22]. Many of these works use spectral graph theory and the graph Laplacian [23] to derive low-dimensional representations by projecting the data on a low-dimensional subspace generated by a small subset of the Laplacian eigenbasis. The graph Laplacian approximates the Laplace-Beltrami operator on a compact manifold [21], [24], in the sense that if the dataset is large and samples uniformly randomly a low-dimensional manifold then the (empirical) graph Laplacian acting on a smooth function on this manifold is a good discrete approximation that converges pointwise and uniformly to the elliptic Laplace-Beltrami operator applied to this function as the number of points goes to infinity [25], [26], [27]. One can go beyond the choice of the graph Laplacian by choosing discrete approximations to other continuous operators and obtaining possibly more desirable spectral bases for the characterization of the geometry of the manifold underlying the data. For example, if the data represents a non-uniform sampling of a continuous manifold, a conjugate to an elliptic Schrödinger-type operator can be used [28], [29], [30].

More in line with the research we developed in this project, several works have proposed multiple transforms for data indexed by graphs. Examples include regression algorithms [31], wavelet decompositions [30], [32], [33], [34], [35], filter banks on graphs [36], [37], de-noising [38], and compression [39]. Some of these transforms focus on distributed processing of data from sensor fields while addressing sampling irregularities due to random sensor placement. Others consider localized processing of signals on graphs in multiresolution fashion by representing data using wavelet-like bases with varying “smoothness” or defining transforms based on node neighborhoods. In the latter case, the graph Laplacian and its eigenbasis are sometimes used to define a spectrum and a Fourier transform of a

signal on a graph. This definition of a Fourier transform was also proposed for use in uncertainty analysis on graphs [40], [41]. This graph Fourier transform is derived from the graph Laplacian and restricted to undirected graphs with real, non-negative edge weights, not extending to data indexed by directed graphs or graphs with negative or complex weights.

The algebraic signal processing (ASP) theory [42], [43], [44], [45] is a formal, algebraic approach to analyze data indexed by special types of line graphs and lattices. The theory uses an algebraic representation of signals and filters as polynomials to derive fundamental signal processing concepts. This framework has been used for discovery of fast computational algorithms for discrete signal transforms [42], [46], [47]. It was extended to multidimensional signals and nearest neighbor graphs [48], [49] and applied in signal compression [50], [51]. The framework proposed that we developed in this project generalizes and extends the ASP to signals on arbitrary graphs.

## B. Overview of our contributions

Our goal was to develop a linear discrete signal processing (DSP) framework and corresponding tools for datasets arising from social, biological, and physical networks. DSP has been very successful in processing time signals (such as speech, communications, radar, or econometric time series), space-dependent signals (images and other multidimensional signals like seismic and hyperspectral data), and time-space signals (video). We refer to data *indexed* by nodes of a graph as a *graph signal* or simply signal and to our approach as *DSP on graphs*  $\text{DSP}_G$ . We developed the basics of linear  $\text{DSP}_G$ , including the notion of a shift on a graph, graph filter structure, graph filtering and graph convolution, graph signal and graph filter spaces and their algebraic structure, the graph Fourier transform, graph frequency, graph spectrum, graph spectral decomposition, and graph impulse and graph frequency responses. With respect to other works, ours is a deterministic framework to signal processing on graphs rather than a statistical approach like graphical models. Our work is an extension and generalization of the traditional DSP, and generalizes the Algebraic Signal Processing theory [42], [43], [44], [45] and its extensions and applications [49], [50], [51]. We emphasize here the contrast between the  $\text{DSP}_G$  and the approach to the graph Fourier transform that takes the graph Laplacian as a point of departure [32], [35], [36], [38], [39], [41]. In the latter case, the Fourier transform on graphs is given by the eigenbasis of the graph Laplacian. However, this definition is not applicable to directed graphs, which often arise in real-world problems, and graphs with negative weights. In general, the graph Laplacian is a second-order operator for signals on a graph, whereas an adjacency matrix is a first-order operator. Deriving a graph Fourier transform from the graph Laplacian is analogous in traditional DSP to restricting signals to be even (like correlation sequences) and Fourier transforms to represent power spectral densities of signals. Instead, we demonstrated that the graph Fourier transform is properly defined through the Jordan normal form and generalized eigenbasis of the adjacency matrix. Finally, we illustrate the  $\text{DSP}_G$  with applications like classification, compression, and linear prediction for datasets that include blogs, customers of a mobile operator, or collected by a network of irregularly placed weather stations, under many other applications.

## Summary of results

$\text{DSP}_G$  extended the algebraic signal processing (ASP) theory introduced in [42],[43],[44], [45],[46] where the shift is the elementary non-trivial filter that generates, under an appropriate notion of shift invariance, all linear shift-invariant filters for a given class of signals. Our key insight in  $\text{DSP}_G$  to build the theory of signal processing on graphs is to identify the shift operator. We adopted the weighted adjacency matrix of the graph as the shift operator and then developed appropriate concepts of z-transform, impulse and frequency response, filtering, convolution, and Fourier transform. In particular, the graph Fourier transform in this framework expands a graph signal into a basis of eigenvectors of the adjacency matrix, and the corresponding spectrum is given by the eigenvalues of the adjacency matrix.

The association of the graph shift with the adjacency matrix in  $\text{DSP}_G$  is natural and has multiple intuitive interpretations. The graph shift is an elementary filter, and its output is a graph signal with the value at vertex  $n$  of the graph given approximately by a weighted linear combination of the input signal values at graph neighbors of  $n$ . With appropriate edge weights, the graph shift can be interpreted as a (minimum mean square) first-order linear predictor. Another interpretation in  $\text{DSP}_G$  of the graph shift comes from Markov chain theory [52], where the adjacency matrix represents the one-step transition probability matrix of the chain governing its dynamics. Finally, the graph shift can also be seen as a stencil approximation of the first-order derivative on the graph.

Because the eigenvalues of the graph shift are in general complex valued, there is an issue that arises in  $\text{DSP}_G$  with defining low and high frequencies, and low-pass, band-pass, and high-pass graph signals or graph filters. In  $\text{DSP}_G$ , we defined low and high frequencies and low-, high-, and band-pass graph signals and filters on generic graphs in a novel way. In traditional discrete signal processing (DSP), these concepts have an intuitive interpretation, since the frequency contents of time series and digital images are described by complex or real sinusoids that oscillate at different rates [33]. The oscillation rates provide a physical notion of “low” and “high” frequencies: low-frequency components oscillate less and high-frequency ones oscillate more. However, these concepts do not have a similar interpretation on graphs, and it was not obvious how to order graph frequencies to describe the low- and high-frequency contents of a graph signal.

In  $\text{DSP}_G$ , we developed an ordering of the graph frequencies that is based on how “oscillatory” the graph spectral components are with respect to the indexing graph, i.e., how much they change from a node to neighboring nodes. To quantify this amount, we introduced the graph total variation function that measures how much signal samples (values of a graph signal at a node) vary in comparison to neighboring samples. This approach is analogous to the approach taken in classical DSP theory, where the oscillations in time and image signals are also quantified by appropriately defined total variations [33]. Once we have an ordering of the frequencies based on the graph total variation function, we define the notions of low and high frequencies, as well as low-, high-, and band-pass graph signals and graph filters.

We applied  $\text{DSP}_G$  in a number of important applications, demonstrating not only its wide applicability as well as the gains of performance it affords when analyzing signals indexed by nodes of a graph. Applications we considered included: signal recovery on graphs, classification, compression, semi-supervised learning, detection of anomalies, and sensor network analysis. For example, signal recovery on graphs recovers one or multiple smooth graph signals from noisy, corrupted, or incomplete measurements. We formulated graph signal recovery as an optimization problem, for which we provided a general solution through the alternating direction methods of multipliers. We showed how signal inpainting, matrix completion, robust principal component analysis, and anomaly detection all relate to graph signal recovery and provided corresponding specific solutions and theoretical analysis. We validated the proposed methods on real-world recovery problems, including online blog classification, bridge condition identification, temperature estimation, recommender system for jokes, and expert opinion combination of online blog classification. On another set of studies, we showed that naturally occurring graph signals, such as measurements of physical quantities collected by sensor networks or labels of objects in a dataset, tend to be low-frequency graph signals, while anomalies in sensor measurements or missing data labels can amplify high-frequency parts of the signals. We demonstrated how these anomalies can be detected using appropriately designed high-pass graph filters, and how unknown parts of graph signals can be recovered with appropriately designed regularization techniques. In particular, our experiments showed that classifiers designed using the graph shift matrix lead to higher classification accuracy than classifiers based on the graph Laplacian matrices, combinatorial or normalized.

The specific framework, theoretical results, analysis methods, and application studies carried out were detailed in papers [1] through [12], see Section C.

## C. References detailing work in the project

- [1] Sandryhaila, A.; Moura, J.M.F., "Nearest-neighbor image model," 19th IEEE International Conference on Image Processing (ICIP), pp.2521-2524, Sept. 30 2012-Oct. 3 2012, doi: 10.1109/ICIP.2012.6467411
- [2] Sandryhaila, A.; Moura, J.M.F., "Discrete Signal Processing on Graphs," IEEE Transactions on Signal Processing, vol.61, no.7, pp.1644-1656, April 1, 2013, doi: 10.1109/TSP.2013.2238935
- [3] Sandryhaila, A.; Moura, J.M.F., "Discrete signal processing on graphs: Graph filters," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.6163-6166, 26-31 May 2013, doi: 10.1109/ICASSP.2013.6638849
- [4] Sandryhaila, A.; Moura, J.M.F., "Discrete signal processing on graphs: Graph Fourier transform," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.6167-6170, 26-31 May 2013, doi: 10.1109/ICASSP.2013.6638850
- [5] Siheng Chen; Sandryhaila, A.; Moura, J.M.F.; Kovacevic, J., "Adaptive graph filtering: Multiresolution classification on graphs," 2013 IEEE Global Conference on Global Conference on Signal and Information Processing (GlobalSIP), pp.427-430, 3-5 Dec. 2013, doi: 10.1109/GlobalSIP.2013.6736906
- [6] Sandryhaila, A.; Moura, J.M.F., "Classification via regularization on graphs," 2013 IEEE Global Conference on Global Conference on Signal and Information Processing (GlobalSIP), pp.495-498, 3-5 Dec. 2013 doi: 10.1109/GlobalSIP.2013.6736923



- [7] Siheng Chen; Sandryhaila, A.; Lederman, G.; Zihao Wang; Moura, J.M.F.; Rizzo, P.; Bielak, J.; Garrett, J.H.; Kovačević, J., "Signal inpainting on graphs via total variation minimization," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) , pp.8267-8271, 4-9 May 2014, doi: 10.1109/ICASSP.2014.6855213
- [8] Sandryhaila, A.; Kar, S.; Moura, J.M.F., "Finite-time distributed consensus through graph filters," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1080-1084, 4-9 May 2014, doi: 10.1109/ICASSP.2014.6853763
- [9] Sandryhaila, A.; Moura, J.M.F., "Discrete Signal Processing on Graphs: Frequency Analysis," IEEE Transactions on Signal Processing, vol.62, no.12, pp.3042-3054, June 15, 2014, doi: 10.1109/TSP.2014.2321121
- [10] Sandryhaila, A.; Moura, J.M.F., "Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure," IEEE Signal Processing Magazine, vol.31, no.5, pp.80-90, Sept. 2014, doi: 10.1109/MSP.2014.2329213
- [11] Siheng Chen; Sandryhaila, A.; Moura, J.M.F.; Kovacevic, J., "Signal denoising on graphs via graph filtering," 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp.872-876, 3-5 Dec. 2014, doi: 10.1109/GlobalSIP.2014.7032244
- [12] Chen, S.; Sandryhaila, A.; Moura, J.; Kovacevic, J., "Signal Recovery on Graphs: Variation Minimization," IEEE Transactions on Signal Processing vol., no., pp., doi: 10.1109/TSP.2015.2441042

## D. References cited

1. C. Chamley, "Rational Herds: Economic Models of Social Learning," 2004, *Cambridge Univ. Press*
2. M. Jackson, "Social and Economic Networks," 2008, *Princeton Univ.*
3. D. Easley and J. Kleinberg, "Networks, Crowds, and Markets: Reasoning About a Highly Connected World," 2010, *Cambridge Univ. Press*
4. M. Newman, "Networks: An Introduction," 2010, *Oxford Univ. Press*
5. J. Whittaker, "Graphical Models in Applied Multivariate Statistics," 1990, *Wiley*
6. S. L. Lauritzen, "Graphical Models," 1996, *Oxford Univ. Press*
7. F. V. Jensen, "Bayesian Networks and Decision Graphs," 2001, *IEEE Comput. Soc. Press*
8. M. I. Jordan, "Graphical models," *Statistical Science (Special Issue on Bayesian Statistics)*, vol. 19, no. 1, pp. 140-155, 2004
9. M. J. Wainwright and M. I. Jordan, "Graphical Models, Exponential Families, and Variational Inference," 2008, *Now Publishers*
10. D. Koller and N. Friedman, "Probabilistic Graphical Models: Principles and Techniques," 2009, *MIT Press*
11. D. Edwards, "Introduction to Graphical Modelling," 2000, *Springer*
12. J. Bang-Jensen and G. Gutin, "Digraphs: Theory, Algorithms and Applications," 2009, *Springer*
13. R. Kindermann and J. L. Snell, "Markov Random Fields and Their Applications," 1980, *Amer. Math. Soc.*
14. A. S. Willsky, "Multiresolution Markov models for signal and image processing," *Proc. IEEE*, vol. 90, no. 8, pp. 1396-1458, 2002
15. J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc.*, vol. 36, no. 2, pp. 192-236, 1974
16. J. M. Hammersley and D. C. Handscomb, "Monte Carlo Methods," 1964, *Chapman & Hall*

17. D. Vats and J. M. F. Moura, "Finding non-overlapping clusters for generalized inference over graphical models," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6368-6381, 2012
18. M. I. Jordan, E. B. Sudderth, M. Wainwright and A. S. Willsky, "Major advances and emerging developments of graphical models," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 17-138, 2010
19. J. F. Tenenbaum, V. Silva and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, 2000
20. S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000
21. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comp.*, vol. 15, no. 6, pp. 1373-1396, 2003
22. D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci.*, vol. 100, no. 10, pp. 5591-5596, 2003
23. F. R. K. Chung, "Spectral Graph Theory," 1996, AMS
24. M. Belkin and P. Niyogi, "Using manifold structure for partially labeled classification," *Proc. NIPS*, p. 929, 2002
25. M. Hein, J. Audibert and U. von Luxburg, "From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians," *Proc. COLT*, pp. 470-485, 2005
26. E. Gine and V. Koltchinskii, "Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results," *IMS Lecture Notes Monograph Series*, vol. 51, pp. 238-259, 2006
27. M. Hein, J. Audibert and U. von Luxburg, "Graph Laplacians and their convergence on random neighborhood graphs," *J. Mach. Learn.*, vol. 8, pp. 1325-1370, 2007
28. R. R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. J. Warner and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Proc. Nat. Acad. Sci.*, vol. 102, no. 21, pp. 7426-7431, 2005
29. R. R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. J. Warner and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods," *Proc. Nat. Acad. Sci.*, vol. 102, no. 21, pp. 7432-7437, 2005
30. R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53-94, 2006
31. C. Guestrin, P. Bodik, R. Thibaux, M. Paskin and S. Madden, "Distributed regression: An efficient framework for modeling sensor network data," *Proc. IPSN*, pp. 1-10, 2004
32. D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann and R. Govindan, "Multiresolution storage and search in sensor networks," *ACM Trans. Storage*, vol. 1, pp. 277-315, 2005
33. R. Wagner, H. Choi, R. G. Baraniuk and V. Delouille, "Distributed wavelet transform for irregular sensor network grids," *Proc. IEEE SSP Workshop*, pp. 1196-1201, 2005
34. R. Wagner, A. Cohen, R. G. Baraniuk, S. Du and D. B. Johnson, "An architecture for distributed wavelet analysis and processing in sensor networks," *Proc. IPSN*, pp. 243-250, 2006
35. D. K. Hammond, P. Vandergheynst and R. Gribonval, "Wavelets on graphs via spectral graph theory," *J. Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129-150, 2011
36. S. K. Narang and A. Ortega, "Local two-channel critically sampled filter-banks on graphs," *Proc. ICIP*, pp. 333-336, 2010

37. S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786-2799, 2012
38. R. Wagner, V. Delouille and R. G. Baraniuk, "Distributed wavelet de-noising for sensor networks," *Proc. CDC*, pp. 373-379, 2006
39. X. Zhu and M. Rabbat, "Approximating signals supported on graphs," *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3921-3924, 2012
40. A. Agaskar and Y. M. Lu, "Uncertainty principles for signals defined on graphs: Bounds and characterizations," *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3493-3496, 2012
41. A. Agaskar and Y. Lu, "Spectral graph uncertainty principle," *IEEE Trans. Inf. Theory*, 2012
42. M. Püschel and J. M. F. Moura, "The algebraic approach to the discrete cosine and sine transforms and their fast algorithms," *SIAM J. Comput.*, vol. 32, no. 5, pp. 1280-1316, 2003
43. M. Püschel and J. M. F. Moura, "Algebraic signal processing theory," [online] Available: <http://arxiv.org/abs/cs.IT/0612077>
44. M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-D time," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3572-3585, 2008
45. M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: 1-D space," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3586-3599, 2008
46. M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Cooley–Tukey type algorithms for DCTs and DSTs," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1502-1521, 2008
47. A. Sandryhaila, J. Kovacevic and M. Püschel, "Algebraic signal processing theory: Cooley–Tukey type algorithms for polynomial transforms based on induction," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 2, pp. 364-384, 2011
48. M. Püschel and M. Rötteler, "Algebraic signal processing theory: 2-D hexagonal spatial lattice," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1506-1521, 2007
49. A. Sandryhaila, J. Kovacevic and M. Püschel, "Algebraic signal processing theory: 1-D nearest-neighbor models," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2247-2259, 2012
50. A. Sandryhaila, S. Saba, M. Püschel and J. Kovacevic, "Efficient compression of QRS complexes using Hermite expansion," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 947-955, 2012
51. A. Sandryhaila and J. M. F. Moura, "Nearest-neighbor image model," *Proc. ICIP*, 2012
52. A. Papoulis and S. U. Pillai, "Probability, Random Variables and Stochastic Processes," 2002, McGraw-Hill
53. S. Mallat, "A Wavelet Tour of Signal Processing, 3<sup>rd</sup> Edition," Academic Press, 2008

1.

**1. Report Type**

Final Report

**Primary Contact E-mail**

Contact email if there is a problem with the report.

moura@ece.cmu.edu

**Primary Contact Phone Number**

Contact phone number if there is a problem with the report

4129832684

**Organization / Institution name**

Carnegie Mellon University

**Grant/Contract Title**

The full title of the funded effort.

An Algebraic Approach to Inference in Complex Networked Structures

**Grant/Contract Number**

AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0087

**Principal Investigator Name**

The full name of the principal investigator on the grant or contract.

Jose M F Moura

**Program Manager**

The AFOSR Program Manager currently assigned to the award

Benjamin Knott

**Reporting Period Start Date**

04/01/2012

**Reporting Period End Date**

03/31/2015

**Abstract**

Analysis and processing of very large data sets, or big data, poses a significant challenge. Massive data sets are collected and studied in numerous domains, from engineering sciences to social networks, biomolecular research, commerce, and security. Extracting valuable information from big data requires innovative approaches that efficiently process large amounts of data and utilize their structure. This research project developed a paradigm for large-scale data analysis based on the discrete signal processing (DSP) on graphs (DSPG). DSPG extends signal processing concepts and methodologies from the classical signal processing theory to data indexed by general graphs. We introduced fundamental concepts of DSPG, including graph signals and graph filters, graph Fourier transform, graph frequency, and spectrum ordering that extended their counterparts from classical signal processing theory. Big data analysis presents several challenges to DSPG, in particular, in filtering and frequency analysis of very large data sets. We showed how to analyze these large data sets by considering product graphs as a graph model that helps extend the application of DSPG methods to large data sets through efficient implementations based on parallelization and vectorization. We illustrated the applicability of DSPG with numerous studies that are relevant in applications.

**Distribution Statement**

This is block 12 on the SF298 form.

### Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

### SF298 Form

Please attach your [SF298](#) form. A blank SF298 can be found [here](#). Please do not password protect or secure the PDF. The maximum file size for an SF298 is 50MB.

[SF 298 FORM.pdf](#)

**Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 50MB.**

[Project-Final-Report-jun30-2015.pdf](#)

[Project-Final-Report-jun30-2015.pdf](#)

**Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.**

### Archival Publications (published) during reporting period:

Sandryhaila, A.; Moura, J.M.F., "Nearest-neighbor image model," 19th IEEE International Conference on Image Processing (ICIP), pp.2521-2524, Sept. 30 2012-Oct. 3 2012, doi: 10.1109/ICIP.2012.6467411

Sandryhaila, A.; Moura, J.M.F., "Discrete Signal Processing on Graphs," IEEE Transactions on Signal Processing, vol.61, no.7, pp.1644-1656, April 1, 2013, doi: 10.1109/TSP.2013.2238935

Sandryhaila, A.; Moura, J.M.F., "Discrete signal processing on graphs: Graph filters," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.6163-6166, 26-31 May 2013, doi: 10.1109/ICASSP.2013.6638849

Sandryhaila, A.; Moura, J.M.F., "Discrete signal processing on graphs: Graph Fourier transform," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.6167-6170, 26-31 May 2013, doi: 10.1109/ICASSP.2013.6638850

Siheng Chen; Sandryhaila, A.; Moura, J.M.F.; Kovacevic, J., "Adaptive graph filtering: Multiresolution classification on graphs," 2013 IEEE Global Conference on Global Conference on Signal and Information Processing (GlobalSIP), pp.427-430, 3-5 Dec. 2013, doi: 10.1109/GlobalSIP.2013.6736906

Sandryhaila, A.; Moura, J.M.F., "Classification via regularization on graphs," 2013 IEEE Global Conference on Global Conference on Signal and Information Processing (GlobalSIP), pp.495-498, 3-5 Dec. 2013 doi: 10.1109/GlobalSIP.2013.6736923

Siheng Chen; Sandryhaila, A.; Lederman, G.; Zihao Wang; Moura, J.M.F.; Rizzo, P.; Bielak, J.; Garrett, J.H.; Kovačević, J., "Signal inpainting on graphs via total variation minimization," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.8267-8271, 4-9 May 2014, doi: 10.1109/ICASSP.2014.6855213

Sandryhaila, A.; Kar, S.; Moura, J.M.F., "Finite-time distributed consensus through graph filters," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1080-1084, 4-9 May 2014, doi: 10.1109/ICASSP.2014.6853763

Sandryhaila, A.; Moura, J.M.F., "Discrete Signal Processing on Graphs: Frequency Analysis," IEEE Transactions on Signal Processing, vol.62, no.12, pp.3042-3054, June 15, 2014, doi: 10.1109/TSP.2014.2321121

Sandryhaila, A.; Moura, J.M.F., "Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure," IEEE Signal Processing Magazine, vol.31, no.5,

pp.80-90, Sept. 2014, doi: 10.1109/MSP.2014.2329213

Siheng Chen; Sandryhaila, A.; Moura, J.M.F.; Kovacevic, J., "Signal denoising on graphs via graph filtering," 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp.872-876, 3-5 Dec. 2014, doi: 10.1109/GlobalSIP.2014.7032244

Chen, S.; Sandryhaila, A.; Moura, J.; Kovacevic, J., "Signal Recovery on Graphs: Variation Minimization," IEEE Transactions on Signal Processing vol., no., pp. ,doi: 10.1109/TSP.2015.2441042

J. Bourgain, F. A. Grünbaum, L. Velázquez, J. Wilkening, "Quantum Recurrence of a Subspace and Operator-Valued Schur Functions," Communication in Mathematical Physics, Digital Object Identifier (DOI) 10.1007/s00220-014-1929-9

**Changes in research objectives (if any):**

None

**Change in AFOSR Program Manager, if any:**

Initially, our PM was Dr. Robert Bonneau. After Dr. Bonneau left AFOSR, our PM is Fr. Benjamin Knott.

**Extensions granted or milestones slipped, if any:**

None.

**AFOSR LRIR Number**

**LRIR Title**

**Reporting Period**

**Laboratory Task Manager**

**Program Officer**

**Research Objectives**

**Technical Summary**

**Funding Summary by Cost Category (by FY, \$K)**

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

**Report Document**

**Report Document - Text Analysis**

**Report Document - Text Analysis**

**Appendix Documents**

**2. Thank You**

**E-mail user**

Jul 02, 2015 13:36:25 Success: Email Sent to: moura@ece.cmu.edu